DOCUMENT RESUME

ED 450 639 HE 033 814

AUTHOR	Goertz, Margaret E.; Duffy, Mark C.
TITLE	Assessment and Accountability Systems in the 50 States, 1999-2000. CPRE Research Report Series.
INSTITUTION	Consortium for Policy Research in Education, Philadelphia, PA.
SPONS AGENCY	National Inst. on Educational Governance, Finance, Policymaking, and Management (ED/OERI), Washington, DC.
REPORT NO	CPRE-RR-046
PUB DATE	2001-03-00
NOTE	47p.; Produced with Kerstin Carlson Le Floch.
CONTRACT	R308A60003
AVAILABLE FROM	Consortium for Policy Research in Education, Graduate School of Education, University of Pennsylvania, 3440 Market St., Suite 560, Philadelphia, PA 19104-3325. Tel: 215-573-0700; Web site: http://www.upenn.edu/gse/cpre.
PUB TYPE	Reports - Research (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Accountability; *Educational Assessment; Elementary Secondary Education; Evaluation Criteria; *School Effectiveness; *State Action; *State Programs; State School District Relationship; State Standards; Student Evaluation

ABSTRACT

Using secondary sources and interviews, this study collected data from the 50 states on how they assess student and school performance and implement accountability. Detailed findings revealed an array of practices by states. The findings answer: (1) "How are states measuring student performance?" (2) "How are states reporting performance on these measures to the general public?" (3) "How are states holding schools, school districts, and students accountable for student outcomes?" (4) "How aligned are accountability policies for Title I and non-Title I schools?" (5) "How are states assisting low-performing schools?" and (6) "What challenges do the federal government and the states face in designing effective and equitable accountability and improvement systems?" (Contains 23 references.) (EV)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 450 639

Assessment and Accountability Systems in the 50 States: 1999-2000

Margaret E. Goertz
Mark C. Duffy

with

Kerstin Carlson Le Floch

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CPRE Research Report Series
RR-046

March 2001

Consortium for Policy Research in Education
University of Pennsylvania
Graduate School of Education

© Copyright 2001 by the Consortium for Policy Research in Education

HE033814

BEST COPY AVAILABLE

Contents

List of Figures and Tables iii
Biographies v
Acknowledgments v
Methodology 2
Measures of Student Performance 2
Characteristics of State Assessment Systems 3
The Role of Local Assessments 7
Assessing Students with Special Needs 8
Setting Student Performance Levels 11
Other Measures of Student Performance 12
Reporting on Student, School, and District Performance 13
Disaggregating and Reporting Disaggregated Test Scores 14
Holding Schools, Districts, and Students Accountable 16
Types of Accountability Systems 16
Performance Measures for School and District Accountability 17
Performance Goals for Schools 19
How Do States Define Progress? 21
Accounting for Subgroup Performance 24
Sticks and Carrots: Consequences for Performance 25
Alignment of General and Title I Accountability Systems Across the States 28
Identifying and Assisting Low-performing Schools 31
Identifying Low-performing Schools 31
Assisting Low-performing Schools 31
Challenges in Implementing Performance-based Accountability Systems 34
Accountability Issues 34
Assessment Issues 35
Equity Concerns 36
Capacity 37
References 39
End Notes 41

List of Figures and Tables

Figure 1. Subjects Included in State Assessment Systems, 1999-2000 3
Figure 2. Grades Tested in Statewide Assessment Systems, 1999-2000 4
Figure 3. State Use of Norm- and Criterion-Referenced Assessments, 1999-2000 7
Figure 4. State Development of Alternate Assessments, 1999-2000 10
Figure 5. Non-cognitive Data Reported at the School and/or District Levels Across the States, 1999-2000 12
Figure 6. Data on Teachers, Resources, and School Climate at the School and/or District Levels, 1999-2000 14
Figure 7. The Disaggregation of Assessment Data Across the States, 1999-2000 15
Figure 8. Types of Assessments Used as Indicators in States with Performance-based Accountability Systems, 1999-2000 18
Figure 9. Number of States Using Non-cognitive Accountability Indicators: 1999-2000 18
Figure 10. The Number of States Using Each of the Three Methods for Defining School Progress, 1999-2000 23
Figure 11. Focus on Accountability: School and District Levels, 1999-2000 27

Table 1. Categories of Defining School Progress in State-defined Performance-based Accountability Systems, 1999-2000 22
Table 2. Alignment of Title I and General State Accountability Systems, 1999-2000 30

Biographies

Margaret Goertz, CPRE's Co-Director for policy and governance, is a Professor in the Graduate School of Education, University of Pennsylvania. Goertz is a policy analyst specializing in the fields of education policy, governance, and finance. She has conducted extensive research on state education reform policies, state teacher testing policies, and state and federal programs for special needs students. Goertz has just completed a multi-year study of standards-based reform in eight states.

Mark Duffy is a researcher with CPRE at the University of Pennsylvania. His primary responsibilities have included coordinating the project, "State and Local Education Policy Reform: From Congress to the Classroom," conducting research and analysis on assessment and accountability systems in the 50 states, and most recently working on an evaluation of New American Schools district services.

Kerstin Carlson Le Floch is a Research Analyst at the American Institutes for Research. Le Floch holds a Ph.D. in Education Organization and Policy Studies from the University of North Carolina at Chapel Hill. Her doctoral research explored the political dynamics preceding the passage of major early childhood legislation in North Carolina and France. Her earlier professional experience also includes two years at the North Carolina Educational Policy Research Center and three years of program development and implementation at the Council for Entrepreneurial Development.

Acknowledgments

This report is based on research supported by a grant (No. OERI-R308A60003) from the National Institute on Educational Governance, Finance, Policymaking, and Management (Office of Educational Research and Improvement, U.S. Department of Education) to the Consortium for Policy Research in Education (CPRE). The authors acknowledge Brenda Turnbull for her helpful comments on an earlier draft. The opinions expressed in this report are those of the authors and do not reflect the views of the National Institute on Educational Governance, Finance, Policymaking, and Management; the U.S. Department of Education; the Office of Educational Research and Improvement; CPRE; or its institutional members.

In recent years, all 50 states have embarked on education initiatives related to high standards and challenging content. A central focus of these efforts has been the establishment of a standards-based reforms have been assessments that measure student performance and accountability systems that are at least partially focused on student outcomes. The policy talk has asserted that these high standards apply to *all* students, and standards-based reform focuses on high achievement for *all* children.

Much of this activity has taken place within the context of the Improving America's Schools Act (IASA) of 1994. This law created major changes in Title I of the Elementary and Secondary Education Act, the $8 billion federal program that provides additional funding to schools with large concentrations of poor children. The legislation was developed in response to concerns about the operation and impact of the Title I program over the previous 25 years: low expectations for educationally disadvantaged students, an instructional emphasis on basic skills, isolation from the regular curriculum, and a focus on procedural compliance rather than academic outcomes (U.S. Department of Education, 1993). Researchers, policymakers, and advocates concluded that a new federal approach was needed to improve education for all students—an approach built on a framework of standards-based reform that is integrated with state and local education reform initiatives.

The provisions of IASA give states a prominent role in Title I. States are expected to establish challenging content and performance standards, implement assessments that measure student performance against these standards, hold schools and school systems accountable for the achievement of all students, and take other steps that promote programmatic flexibility and foster instructional and curricular reform. States are also expected to align their Title I programs with these policies to ensure that disadvantaged students are held to the same high standards. In addition, the 1997 revisions to the Individuals with Disabilities Education Act (IDEA) require states to include students having disabilities in state and district assessment programs, with appropriate accommodations, and to disaggregate and report their test scores.

This report uses data collected from the 50 states to describe state assessment and accountability systems and to examine the extent to which state policies meet the objectives of federal policy. Specifically, we look at:

- How are states measuring student performance?
- How are states reporting performance on these measures to the general public?
- How are states holding schools, school districts, and students accountable for student outcomes?
- How aligned are accountability policies for Title I and non-Title I schools?
- How are states assisting low-performing schools?
- What challenges do the federal government and the states face in designing effective and equitable accountability and improvement systems?

Methodology

The findings reported here are drawn from a 50-state survey of state assessment and accountability systems conducted by the Consortium for Policy Research in Education (CPRE) between February and June 2000. We concentrated on collecting state policies that were in place during the 1999-2000 school year. We used a four-step process to collect and verify our data. First, we collected and analyzed existing data from secondary sources: weekly and special issues of *Education Week* such as *Quality Counts* (1999, 2000), the Council of Chief State School Officers (2000), the American Federation of Teachers (1998, 1999), and state department of education web sites. We then conducted semi-structured interviews with directors of assessment, accountability, and Title I programs in each state to confirm, clarify, and update information from written sources. We also used these interviews to identify proposed changes in state policies. Materials provided by the respondents often supplemented these interviews. Third, we drafted extensive profiles that described each state's policies and practices on assessment, inclusion, reporting, accountability, assistance, and Title I. Finally, we asked state respondents to verify the written profiles, and we incorporated suggested changes and corrections into the final profiles. The state profiles are available on the CPRE web site: www.gse.upenn.edu/cpre/.

The information presented in this report was current when the profiles were verified by each state, generally between April and July 2000. One of the challenges in conducting this study was the transitory nature of many state accountability systems. Several states were in the process of redesigning assessment and accountability systems to meet state or federal policy requirements, including those of Title I. Even states with established accountability systems, such as Kentucky, have modified their policies in response to technical or political concerns. Many states were putting new assessment or accountability systems in place during 2000-2001 and other states will start implementing new policies in 2001 or later.

Thus, we found ourselves studying a moving target. We addressed this policy flux in the following way. The data reported here represent policies in place in 1999-2000 *unless* a state: had enacted and planned to implement revised policies in 2000-2001, had enacted new policies for 2000-2001 and reported it was awaiting federal Title I approval of their new system, or had proposed new policies for the 2000-2001 school year and was awaiting approval by its state board of education. In these three cases, we treated the new policies as current practice. If a state had enacted or proposed policies scheduled to be implemented *after* the 2000-2001 school year, we reported the policies in place in 1999-2000 as current practice.

Measures of Student Performance

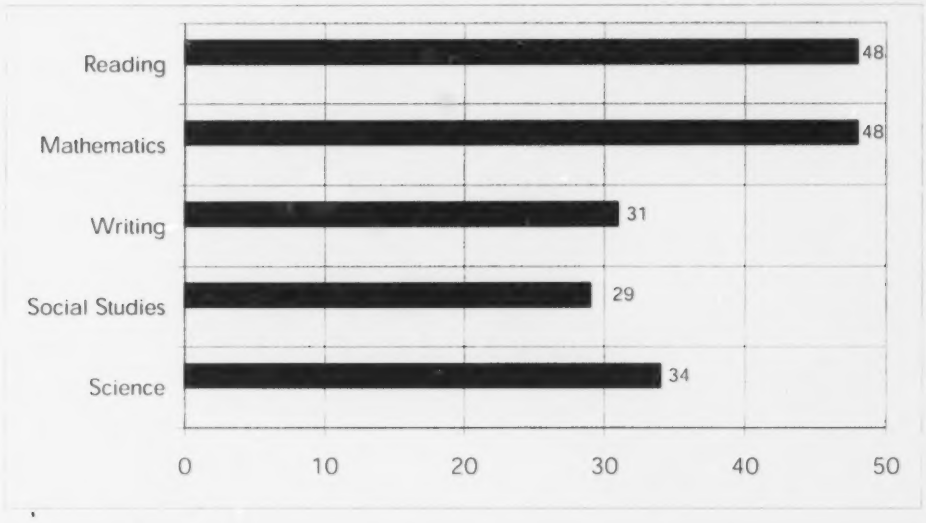
The state accountability systems that have emerged over the last decade focus on student outcomes, particularly student performance on state assessments. Some states also collect measures of non-cognitive performance, such as attendance and dropout rates, and a few states incorporate local assessment results in their accountability measures. This section describes how states measure student outcomes, paying particular attention to the characteristics of state assessment systems. How states report these outcomes and use these measures in accountability systems will be discussed in subsequent sections of the report.

Characteristics of State Assessment Systems

Forty-eight states use a state assessment as the principal indicator of school performance. The other two states—Iowa and Nebraska—require their districts to test students in specified grades or grade spans, but leave the choice of assessment instrument to the locality. States differ, however, in the subjects and grades they assess and the types of tests they administer.

What subjects and grades are tested? All 48 states with statewide assessment systems test students in mathematics and English/language arts or reading (Figure 1). Fewer states test writing (31), science (34), and social studies (29).

Figure 1. Subjects Included in State Assessment Systems, 1999-2000

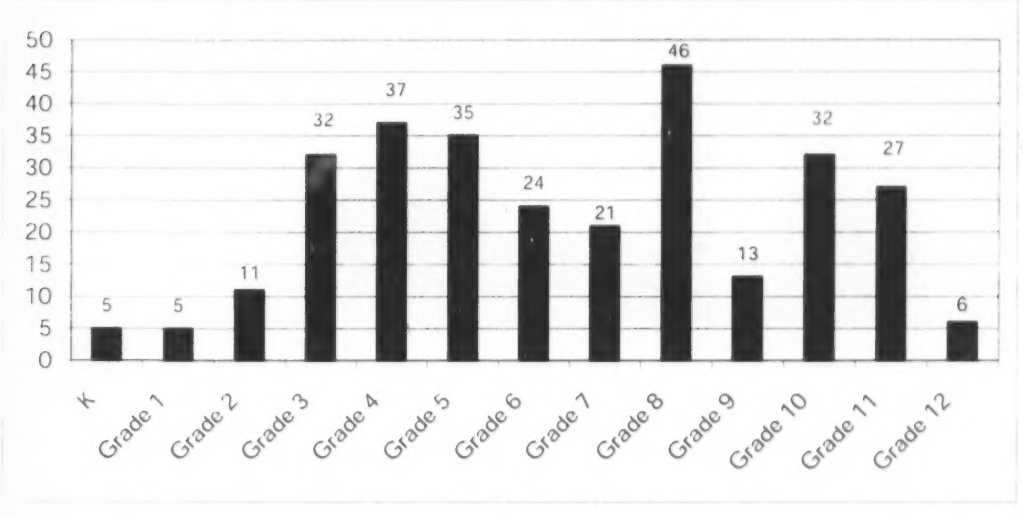


The Improving America's Schools Act (IASA) of 1994 requires that states test students at least once during each of three grade spans: third-to-fifth, sixth-to-ninth, and tenth-to-twelfth. But, states assess students considerably more often, with some states testing students in almost every grade. The 48 states fell into one of three categories in 1999-2000:

- testing students in one grade per subject at the elementary school, middle school, and high school levels;
- testing the same subject, using the same assessment, in consecutive grades between the second or third grade and at least the eighth grade; or
- testing consecutive grades between second or third grade and eighth grade in different subjects or using multiple assessments.

The majority of the 48 states tested students at specific points in each of the three grade spans, similar to the federal requirements. Maine, Oregon, and Michigan are examples of such states. Maine students are tested only in the third, eighth, and eleventh grades. In Oregon, students in the third, fifth, eighth, and tenth grades participate in state assessments. The Michigan Educational Assessment Program tests reading and mathematics in the fourth and seventh grades, and writing, social studies, and science in the fifth and eighth grades; the High School Proficiency Test assesses all five subjects in the eleventh grade. States concentrate testing activity in the third, fourth, fifth, eighth, and tenth grades (Figure 2).

Figure 2. Grades Tested in Statewide Assessment Systems, 1999-2000



In 1999-2000, twelve states tested students in the same subject areas in consecutive grades: Alabama, Arizona, California, Florida, Idaho, Mississippi, New Mexico, North Carolina, South Carolina, Tennessee, Texas, and West Virginia. Arizona and Florida use a norm-referenced test in all grades, but their state standards-based assessments in only four grades.

A few more states test in every grade, but in differing subject areas or using different tests. For example, Kentucky's Commonwealth Accountability Testing System (CATS) assesses reading, writing, and science in the fourth and seventh grades, and mathematics and social studies in the fifth and eighth grades. The assessment of these subjects is spread across the high school years: reading in the tenth grade; mathematics, science, and social studies in the eleventh grade; and writing in the twelfth grade. Students in grades not tested by CATS (third, sixth, and ninth) take the Comprehensive Test of Basic Skills (CTBS) in reading and mathematics. Louisiana and Maryland test third-to-ninth grade students in the same subjects every year, but use different assessments. Louisiana tests students in the third, fifth, sixth, and seventh grades with the Iowa Tests of Basic Skills (ITBS), in the fourth and eighth grades with the criterion-referenced Louisiana Educational Assessment Program for the 21st Century (LEAP 21), and in the ninth grade with the Iowa Tests of Educational Development (ITED). Maryland administers its

performance-based assessment, the Maryland State Performance Assessment Program (MSPAP), in the third, fifth, and eighth grades, and the CTBS in the alternating second, fourth, and sixth grades.

Testing before the third grade. Most state assessment systems begin in the third grade. Several states, however, are developing assessments for earlier grades as a means of identifying and diagnosing learning problems early in a student's educational career. Seven states—Alabama, Georgia, Idaho, South Carolina, Vermont, Washington, and West Virginia—test students in kindergarten, first, or second grade. Some of these assessments, like the Alabama Reading Assessments or the Vermont Developmental Reading Assessment, are designed to measure reading ability in the first or second grade. Four of these states administer more general diagnostic or elementary readiness assessments to determine the development of basic skills in the early grades. Georgia and West Virginia administer diagnostic tests in kindergarten.

Other states, such as Connecticut, Kansas, New Mexico, Oklahoma, and Texas, require districts to test reading in early grades, but allow districts to choose the assessment instrument. Texas requires every district to administer a reading instrument to each student in kindergarten through the second grade. Local districts can select an assessment from the Commissioner's list of acceptable tests, or adopt another instrument with the approval of the district committee.

The increase in high school testing. In 1996-1997, eighteen states required students to pass a state-administered test in order to graduate from high school (Bond, Roeber, and Connealy, 1998). By 2008, high school students in 28 states will have to meet this requirement. Two states will require that students either pass the state or a local high school assessment. In another seven states, student performance on a state assessment may be noted on a student's transcript or diploma, but passing a state test is not required for high school graduation. For example, results from the Connecticut Academic Performance Test (CAPT), which is administered in tenth grade, are reported on student transcripts. High school students are awarded a Certificate of Mastery in each subject area they score at or above the state performance goals. In Arkansas, student scores on end-of-course examinations also become part of student transcripts and permanent school records. Finally, some states like Kentucky and Vermont use the results of high school assessments in their accountability systems, but the tests have no consequences for students.

Most state high school tests assess a student's general knowledge of English/language arts and mathematics, and often of science and social studies as well. Ten states, however, have or are developing end-of-course examinations for their high school students. Five of the ten states—Maryland, Mississippi, New York, Tennessee, and Virginia—will require students to pass end-of-course examinations to graduate from high school. New York has been phasing in a requirement that high school students pass new versions of that state's Regents examinations in English, mathematics, global history and geography, United States history and government, and science (as chosen by the student from chemistry, physics, earth science, and biology). Tennessee students who enter high school in 2001-2002 must pass new end-of-course examinations in Algebra I, English II and biology.[1] In all five of these states, the end-of-course assessments are replacing basic skills exit exams.

Four states—Arkansas, North Carolina, Oklahoma, and South Carolina—record the results of end-of-course examinations on student transcripts, but do not use these assessments as exit tests. Students in North Carolina and South Carolina must pass a separate competency test to graduate from high school. Finally, Texas students have had the option of passing either the state's end-of-course examinations or competency test as a high school graduation requirement. The end-of-course examinations will be phased out, however, when Texas implements its new eleventh grade exit test which will cover the four core subjects of language arts, mathematics, social studies, and science.

Norm- or criterion-referenced tests? As states have developed assessment systems to measure student progress toward state standards, some educators and researchers have questioned whether and how well norm-referenced tests are aligned with state standards, and whether the tests are appropriate measures of student performance on challenging standards. Norm-referenced tests measure the knowledge and skills of students across the country, while criterion-referenced tests measure knowledge and skills that are specific to a state (or district). We were interested, therefore, in knowing what kinds of tests the states included in their assessment systems.

Seventeen states administer only criterion-referenced tests and two states—Montana and South Dakota—use only norm-referenced tests in their state assessment systems (Figure 3). The remaining 29 states administer a combination of criterion-referenced and norm-referenced tests. These states differ, however, in how they use the two kinds of tests.

- Four states—Alabama, Idaho, West Virginia, and Wisconsin—use a norm-referenced test as their principal assessment instruments, but administer criterion-referenced tests in a limited number of subjects and grades. Wisconsin supplements its norm-referenced assessment with a criterion-referenced reading assessment in the third grade. Idaho and West Virginia administer writing tests in addition to the ITBS (Idaho) or Stanford Achievement Test (SAT) (West Virginia). Students in Alabama take the SAT-9 test in the third through eleventh grades but must pass a criterion-referenced test to graduate from high school. Illinois, in contrast, uses a criterion-referenced test to assess students in most grades, but relies primarily on norm-referenced items for its eleventh grade test.

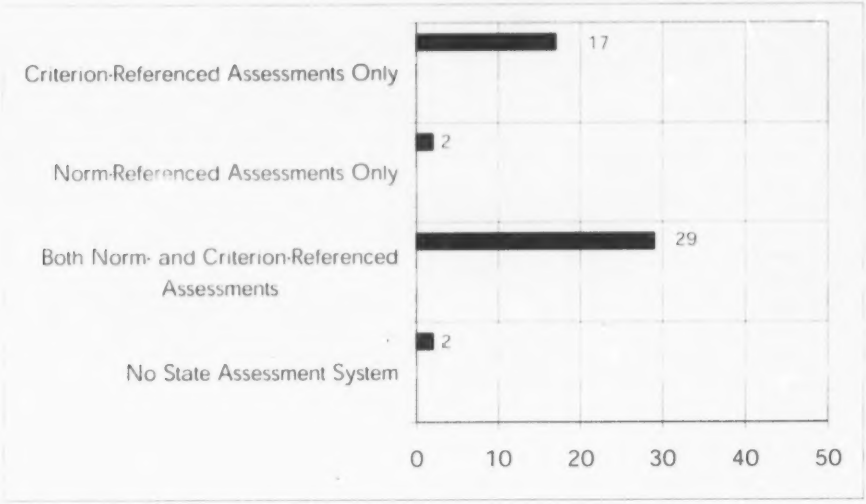
- Ten of the states administer norm-referenced and criterion-referenced examinations in different grades. As noted earlier, Maryland and Kentucky use norm-referenced tests in grades not included in their criterion-referenced state assessments.

- Eight other states give both kinds of assessments in the same grades. Arizona students in the second through the eleventh grades take the SAT-9; students in the third, fifth, eighth, and tenth grades also take the criterion-referenced Arizona Instrument to Measure Standards (AIMS). Mississippi tests students with the state's criterion-referenced test in the second through eighth grades and the Terra Nova in the third through eighth grades.

- Finally, six states—California, Delaware, Indiana, Missouri, New Mexico, and Tennessee—have developed state tests to produce both nationally-normed results and measures of student performance on each state's standards. These states have generally worked with a test publisher to develop an assessment that combines items from a national, norm-referenced test

such as the SAT-9 or the TerraNova, with customized items that cover state standards not measured by the norm-referenced questions.

Figure 3. State Use of Norm- and Criterion-Referenced Assessments, 1999-2000



State policymakers appear to have three reasons for including norm-referenced tests in their assessment systems. First, parents, policymakers, and the public want a way of comparing the performance of their students to students outside their states. The National Assessment of Educational Progress (NAEP) provides this kind of comparison, but only at the state level, periodically, and for a limited number of subjects. In addition, NAEP is not a high-profile assessment like national commercial tests. Second, perhaps for the first reason, some state legislatures require the administration of norm-referenced assessments. Finally, the cost of developing state-based criterion-referenced tests is too high for small states.

The Role of Local Assessments

The expansion of state testing programs has reduced the role of local tests in state assessment and accountability systems. Only roughly one in five states has local testing requirements.

The two states without statewide assessment systems—Iowa and Nebraska—require local districts to assess students, with a test of their choosing, in particular subjects and in specific grade spans. Nebraska, however, will implement a statewide writing test in the spring of 2001. Nebraska is also phasing in tests in reading, mathematics, science, and history/social studies, but districts will choose from four local tests selected by the state or bring their own tests up to the models' standards.

Three other states have taken steps to incorporate local assessment results into their student or school accountability measures. Colorado school districts, to comply with the state's Basic Literacy Act, must use evidence from "individual reading assessments" and the state assessment

to determine the reading proficiency of third grade students. Districts may use other reading assessments as part of the "body of evidence" they use to determine whether students can be promoted to fourth grade reading. Maine gives districts the option of using the new Maine Assessment Portfolio system to supplement information generated by the Maine Educational Assessment. Vermont's new accountability system gives greater weight to state assessments, but schools are encouraged to use other assessments and, if approved by the State Board of Education, they may select one or more local assessment for accountability purposes. The state will determine the individual and combined maximum weight of the local assessments (relative to the state assessments) in the accountability system. Local assessments could count as much as 30 percent of a district's accountability measure.

Finally, several states require districts to assess early literacy skills as a means of identifying students who need help in reading in the primary grades.

Few states require local districts to administer additional tests, but a more intensive study of eight states showed that districts often supplemented state tests with local assessments so they could track student progress during the year (Goertz, 2000; Goertz, Massell, and Chun, 1998).[2] Two of the three Kentucky study districts added district testing: one district administered the CAT-5 in most grades; another administered school-selected reading and mathematics tests three times a year. Two of the Maryland study districts used district-designed end-of-unit tests and teacher-generated running records to measure student progress against district goals (which are aligned with state goals). Both of the California districts in the study supplemented the state's SAT-9 test with performance-based assessments. One California district, which has an extensive and innovative testing program and a commitment to multiple measures of student achievement, assessed literacy through running records, a benchmark book program for early literacy, performance writing tasks, and standardized district-purchased tests for reading. Mathematics was assessed by a homegrown facts test, performance assessments, and standardized tests. The Texas study districts were less likely to administer additional local tests, possibly because the state assessment covers nearly all grades. One district, however, developed district tests to measure student performance on its more rigorous and comprehensive district standards.

These local assessments varied in their alignment with the state assessment systems. The district assessments used in the three Maryland districts worked well in supplementing the state assessment, while the assessments used to supplement the CATS in Kentucky were not so well aligned. An external analysis of Michigan's state assessments found that they "embody a more comprehensive and demanding set of expectations for Michigan's students than might be assumed from reading the state standards alone," leaving district assessment directors at a loss about the appropriateness of local testing (Achieve, Inc., 1998). In some states, local assessments were not intended to be aligned, but to generate individual student results or to measure the "value added" by the local school district, as was the case in one Texas district.

Assessing Students with Special Needs

States face both technical and political challenges as they try to include students with disabilities and students with limited-English proficiency in their assessment systems. There are two major reasons to include special student populations in assessment and accountability systems. The first

reason is to improve the quality of educational opportunities afforded special needs students. Holding educators accountable for test scores, the theory goes, will increase these students' access to a high-quality, standards-based general education curriculum. The second reason is to provide useful information about the performance of special needs students so parents and the public know how well the school is educating all the children.

Achieving these goals, however, requires that districts assess special needs students on the content of the standards-based curriculum, disaggregate and report their scores, and include their scores in school or district accountability measures. The provisions of Title I and the Individuals with Disabilities Education Act are intended to meet these requirements. States must include all students in the grades they test, and must assess all students against the same content and performance standards. If standard assessment procedures cannot provide this information for students with diverse learning needs, such as students with disabilities or English-language learners, states must make reasonable test adaptations and accommodations, or provide alternate assessments. These changes, however, must yield accurate and reliable information on student mastery of the content covered by state standards. Districts must disaggregate test results if the data are statistically sound, and report the results with the same frequency as results reported for the general population.

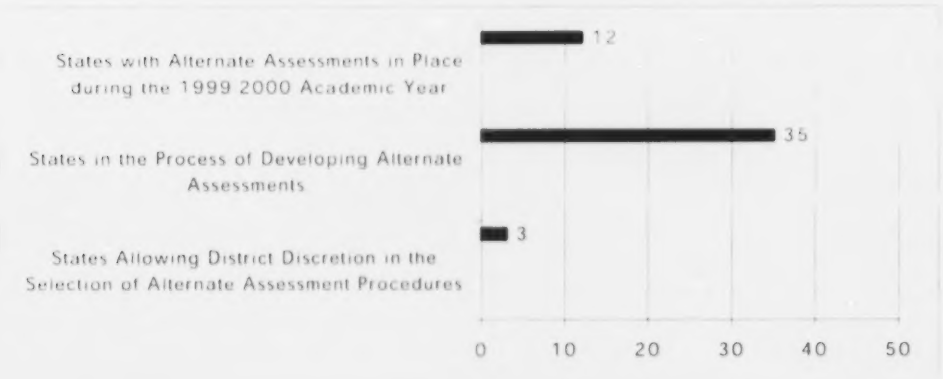
In addressing the requirements of Title I and IDEA, state policymakers face a seemingly intractable problem: How can they include all students in state assessment systems while ensuring that these assessments generate valid data? Can we assess all students with instruments and under conditions that yield construct-relevant information and generate valid inferences about their knowledge and performance? These issues of test validity and construct-relevance underlie the decisions policymakers make about: who gets tested on what and how, whose test scores are reported and how, and whose scores are included in accountability measures. This section looks at who gets tested and how. We examine the issues of reporting and accountability later in the report.

According to the National Center on Educational Outcomes (1999), "more students with disabilities are participating in statewide testing" and "most states are in the process of developing alternate assessments." The research results discussed in this report support these findings. States report testing more students with disabilities and offering a range of test accommodations and modifications.[3] States appear to offer a broader range of accommodations and modifications to their own criterion-referenced assessments. When using commercial, norm-referenced tests, however, states may be limited in the accommodations allowed by the test publisher. Twelve states had an alternate assessment in place during 1999-2000. Thirty-five states were in the process of developing alternate assessments; 26 of these states planned to have their assessments in place for the 2000-2001 school year (Figure 4).

Only three states—Florida, Montana, and Ohio—plan to let districts select alternate assessments or assessment procedures. Ohio has developed a model set of alternative assessment procedures districts may use, or districts may select different alternative measures. The Florida Department of Education has provided guidelines for "critical components" of alternate assessments, and has instructed district IEP (individual education plan) committees to indicate the alternate assessment

strategy appropriate for each student based on the standards, annual goals, short-term objectives, curriculum, and instruction.

Figure 4. State Development of Alternate Assessments, 1999-2000



States face ongoing challenges in determining student eligibility for alternate assessments, aligning these tests with state standards, and scoring and reporting test results (Sack, 2000). States report that they are monitoring exclusion rates of students with disabilities; some states are incorporating exclusion rates into their school accountability measures.[4]

The story is different for students with limited proficiency in English, or English-language learners. Tests administered in English to students with limited-English proficiency can be more of an assessment of English ability than content knowledge (President's Advisory Commission on Educational Excellence for Hispanic Americans, 2000). Therefore, states have developed a variety of policies regarding whether and when English-language learners are included in state assessments. A first group of states exempts English-language learners based on the length of their residency. California, for example, requires all English-language learners to take the state assessment, but allows students who have been in the public school system less than one year to take a Spanish language assessment (the SABE 2) as well. Other states exclude students who have resided in the United States or in their state up to three years, if they are enrolled in a bilingual or English-as-a-Second-Language (ESL) program.

A second group of states exempts students based on the length of time spent in an ESL or bilingual education program. Florida excludes students with less than two years of ESL; students with two or more years in ESL must be tested in English, but can have accommodations such as additional time or dividing the test into shorter periods. Many other states offer similar test accommodations to English-language learners. A third set of states exempts students based on their level of English proficiency. English-language learners in Nevada must pass the Language Acquisition Skills assessment in order to be included in the state assessment. Colorado exempts non-English speaking students who score at the first or second levels on a five-stage language proficiency rubric. Texas exempts Spanish-speaking students, based on their level of English proficiency, from its regular third-through-eighth-grade testing program, but requires all students to take the tenth grade exit test in English.

Finally, 15 states offer versions of some of their state assessments in languages other than English. Arizona, New Mexico, and Texas offered Spanish versions of some state tests. New York provides mathematics tests in four languages and will translate into five languages high school examinations in subjects other than English. Utah limits its second-language assessments to the state's pre-kindergarten test. Providing assessments in a student's native language has become especially difficult in areas of the country with large Native American populations and multiple dialects. The President's Advisory Commission on Educational Excellence for Hispanic Americans (2000) has raised concerns, however, about the rigor of Spanish-language and translated versions of state assessments and about the suitability of many test accommodations for English-language learners.

Setting Student Performance Levels

Title I requires states to establish at least three levels of student performance on state assessments—advanced, proficient, and partially proficient—to show how well students are mastering the state content standards.

Nearly all of the states with statewide assessments reported they had student performance levels in place for the 2000-2001 school year. A few states that use norm-referenced tests had not developed performance standards for these assessments, reporting results only by national percentile rank. As of January 2001, the U.S. Department of Education had approved performance standards in only 28 states (U.S. Department of Education, Planning and Evaluation Service, 2001).

Student Performance Levels: Two Approaches

Colorado and Minnesota provide two approaches to setting student performance levels.

Colorado uses four proficiency levels to describe student performance:

- Advanced
- Proficient
- Partially proficient
- Unsatisfactory

Minnesota describes four levels of student performance on third and fifth grade tests:

- Level 1: Little evidence of knowledge and skills
- Level 2: Partial knowledge and skills
- Level 3: Solid academic performance
- Level 4: Superior performance

Minnesota has only one performance level for its eighth and tenth grade reading and mathematics tests—correctly answering 75 percent of the questions.

Most states (37) have created four-to-five student performance levels, generally adding an additional category of partial proficiency. Kentucky expanded its performance reporting to eight levels of achievement in order to capture progress below proficiency. Two states, however, have only two proficiency categories. Some states set different performance levels for different tests. Michigan has designed a separate system for each subject assessed, with two to four proficiency categories, depending on subject and grade level.

States use performance categories for multiple purposes, including student reporting, retention, and awarding high school diplomas. Setting cut scores for each performance level can be difficult and controversial. New Jersey recently considered changing how it translated the collective raw scores for the fourth grade writing

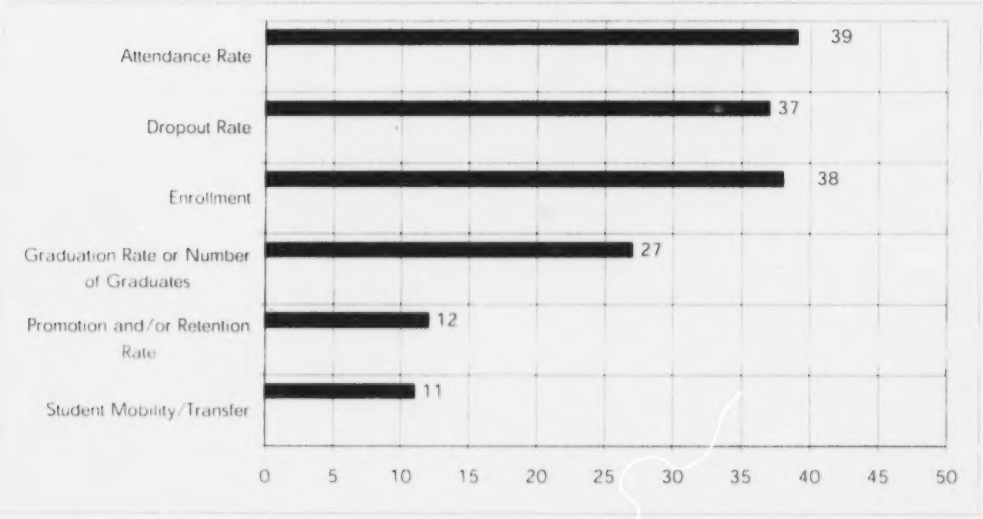
portion of the state's language arts test into three scoring categories because a high percentage of students scored in the lowest category (Avril, 2000). To minimize such problems, many states use the piloting process not only to examine the validity and reliability of an assessment, but also to determine how well the state's students will perform at different cut points.

The language used to describe student performance is not standardized across states. The terminology used to describe performance that meets standards includes *proficient* (in Colorado), *solid academic performance* (Minnesota), and *meets standards* (Michigan). Students performing below standards are considered *partially proficient*, having *partial knowledge or little evidence of knowledge and skills*, and *pre-emerging* in different states. The proficiency labels garnered widespread public attention in Michigan because high school assessment performance is noted on students' high school diplomas. In 1997, after parental protests, the Michigan legislature suspended endorsing diplomas with *novice* and *not yet novice* designations. In 1998, legislative amendments changed the name and proficiency levels of the tests; the two lowest levels are now called *endorsed at basic level* and *unendorsed.*

Other Measures of Student Performance

States often collect measures of student performance other than tested achievement. As shown in Figure 5, local report cards in more than half of the states include attendance rates or average daily attendance (39), dropout rates (37), graduation rates (27), and enrollment (38). More than one-fifth of the states also include student mobility (11), and promotion or retention rates (12) on the report cards. About one-half of the states report information on Advanced Placement course-taking and test scores and average SAT or ACT scores.

Figure 5. Non-cognitive Data Reported at the School and/or District Levels Across the States, 1999-2000



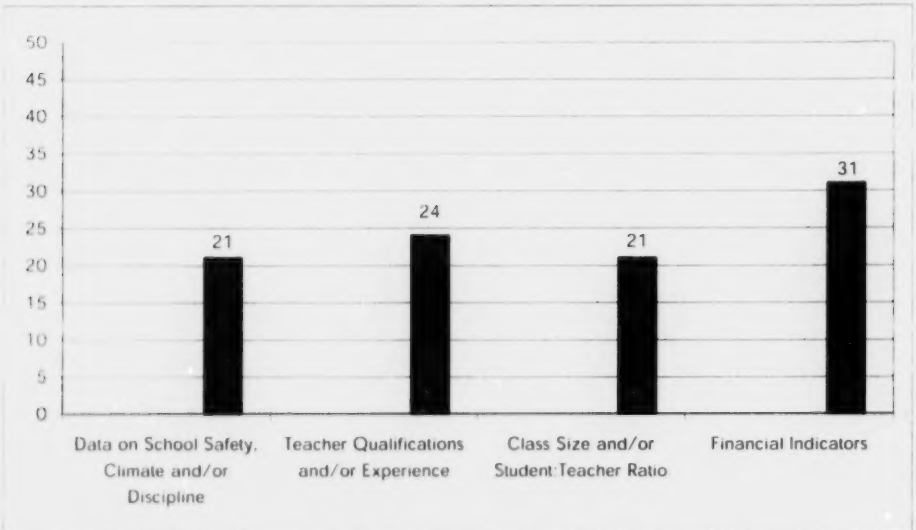
Reporting on Student, School, and District Performance

Public reporting is the most basic form of accountability. Schools give an account of their programs and performance. The public can then use this information to demand improvements in their schools, or possibly to choose alternative schools for their children. All states and districts receiving Title I funds must issue annual school, district, and state report cards that include information on student achievement.

All 50 states currently produce or require local school districts to publish district or school report cards. School report cards are prepared in 40 states (Council of Chief State School Officers, 2000). The report cards include, at a minimum, student performance on state or local assessments. States that have established student performance levels, such as advanced, proficient, basic, and below basic, generally report the percentage of students scoring at each of these levels. If the state assessment is norm-referenced or includes a norm-referenced component, scores are also usually reported as scale scores or by the national percentile ranking. In some cases, however, scores from norm-referenced assessments are used to place students into proficiency levels. In West Virginia, for instance, students are listed as advanced, proficient, basic, or minimal performance in terms of TerraNova scale scores. Assessment scores are further disaggregated by grade, subject, and to a lesser extent, by other demographic categories. States that have created school or district performance categories also report that information on school or district report cards.

Public reports also include data on non-cognitive measures of student performance and other factors influencing education. As noted earlier, nearly 40 states report student attendance, dropout rates, and graduation rates. Many states include indicators of school climate, teacher quality, and fiscal resources (Figure 6). The most commonly reported input measures include data on school discipline, safety, and climate (21 states); teacher qualifications and experience (24 states); class size or student-to-teacher ratio (21 states); and financial information, such as per-pupil expenditure and school and district revenues (31 states). School safety, climate, and discipline data include the number of violent incidents against teachers and students, and the number of student suspensions and expulsions. In addition to information on class size or student-to-teacher ratios, schools and districts also commonly report the number of staff in a school and the ratio of administrators to students. Data on teacher experience and qualifications generally include the number or percentage of teachers with graduate degrees and number of years of teaching experience.

Figure 6. Data on Teachers, Resources, and School Climate at the School and/or District Levels, 1999-2000

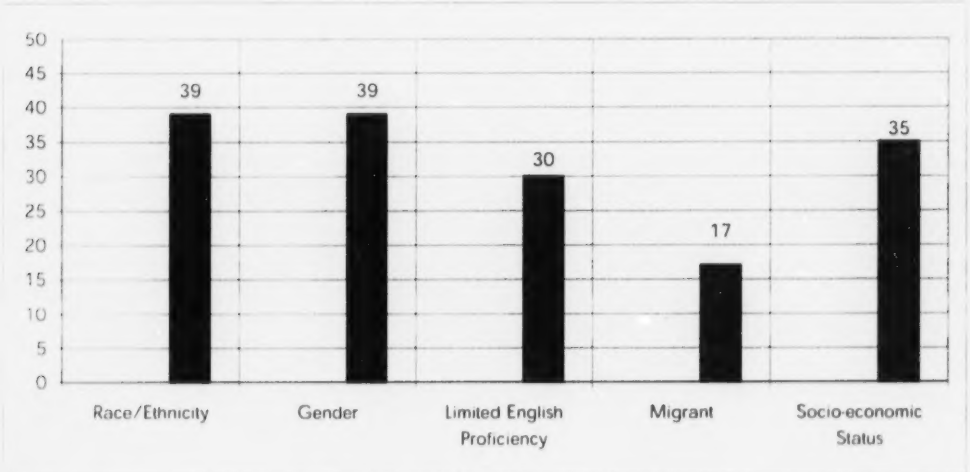


Disaggregating and Reporting Disaggregated Test Scores

The Title I legislation requires that states *enable* the disaggregation of state test scores at the school, district, and state levels by six student groups: gender, major racial/ethnic group, English-proficiency status, disabled vs. non-disabled status, migrant status, and economically disadvantaged vs. non-economically disadvantaged status (Sec.1111(b)(3)(I)). The U.S. Department of Education further requires that states provide for reporting the disaggregated data if they are statistically sound (U.S. Department of Education, 1999). In addition, the 1997 amendments to the Individuals with Disabilities Education Act require that the performance of students with disabilities is disaggregated from the scores of other students and is reported in the same way as the performance of other students.

Thirty-nine of the states with statewide assessment systems report that they disaggregate test data by race/ethnicity and gender (Figure 7). Student data are disaggregated by socio-economic status (usually free or reduced-lunch eligibility) in 35 states, by English- proficiency status in 30 states, and by migrant status in 17 states. These states, however, do not necessarily report disaggregated test scores to the public.

Figure 7. The Disaggregation of Assessment Data Across the States, 1999-2000



In the latest survey by the National Center on Educational Outcomes (1999), all special education directors in states having statewide assessments reported that the scores of test-takers receiving special education services were disaggregated. In 1997, twenty-two states did not disaggregate these data. It is not known, however, how many states publicly report the performance of students with disabilities. Our data show that states take one of the five following approaches in disaggregating and reporting scores of students with disabilities and of English-language learners:

1. The states neither disaggregate nor report these scores.
2. The states disaggregate but do not publicly report the scores.
3. The states do not disaggregate but include the scores in aggregate score reports.
4. The states report the scores of tests taken under standard conditions or under conditions that do not interfere with the comparability of scores of students tested under regular conditions.
5. The states disaggregate and report all scores.

Delaware is an example of a state in the fourth category. Students are assessed under one of five testing conditions: regular conditions, with accommodations that do not interfere with the comparability of their scores to scores of students tested under regular conditions, with accommodations that interfere with comparability, an alternative portfolio assessment, or exemption for limited-English proficiency (one time only, and if in Delaware schools for less than two consecutive years). Only scores of tests taken under the first two conditions are included in school, district, and state score reports.

Arizona disaggregates and reports the scores of all test-takers by category of testing condition: all students—standard conditions, regular education—standard conditions, special education—standard conditions, and special education—non-standard conditions. Indiana takes a similar approach, reporting scores as: all tested, general education with and without accommodations, and special education with and without accommodations.

Most states include the scores of English-language learning students in their aggregate reports, but do not necessarily report these scores separately.

Holding Schools, Districts, and Students Accountable

State accountability systems create incentives for students, schools, and school districts to focus on student achievement and continuous progress. The type and strength of these incentives are determined largely by the design of the accountability system, particularly who sets what goals for the system, the measures of adequate progress, and the consequences of meeting (or not meeting) these goals. State accountability systems vary along all of these dimensions.

Types of Accountability Systems

We have grouped the state accountability systems into three categories based on who sets goals for the system and the extent to which schools and districts are held accountable for student performance. The three categories are: public reporting systems, locally-defined accountability systems, and state-defined accountability systems.

Public reporting. As discussed in the preceding section of this report, all states report annually on student performance. Only 13 states, however, use public reporting as their primary accountability mechanism.[5] In most of these states, districts are required to administer and report the results of a statewide assessment. The two states without statewide assessment systems require districts to administer tests of their choosing. Districts must provide performance reports to their communities that include student achievement and possibly other measures of student performance. None of these states rank or rate school or district performance, or identify low-performing schools. These tasks are left to the public and local educators. A few of these states do hold students accountable for academic performance through high school exit examinations.

Three of the states that rely on public reporting—Alaska, Georgia, and Hawaii—are developing school-based accountability systems that are supposed to go into effect in 2001-2002 or later. Policymakers in other states, such as New Hampshire and Minnesota, have proposed such systems, but have not yet gained the political support needed to gain passage of accountability systems in their state legislatures.

Locally-defined accountability systems. A few other states have developed accountability systems that emphasize local standards and local planning. These states allow districts to establish criteria for school performance, but use strategic plans or district and school improvement plans to hold districts accountable for student performance. For example, under the Kansas Quality Performance Accreditation system, schools establish student exit outcomes and target weak areas for improvement. Schools are expected to show evidence of continuous student improvement as part of their accreditation review. Wyoming holds districts accountable through an accreditation process that requires periodic district site visits and all schools and districts to submit an accreditation packet to the state department of education. This continuous improvement process involves the following elements: standards, measures, a school

improvement planning profile, school improvement goals, an action plan, targets for at-risk subgroups, staff development, and school improvement results.

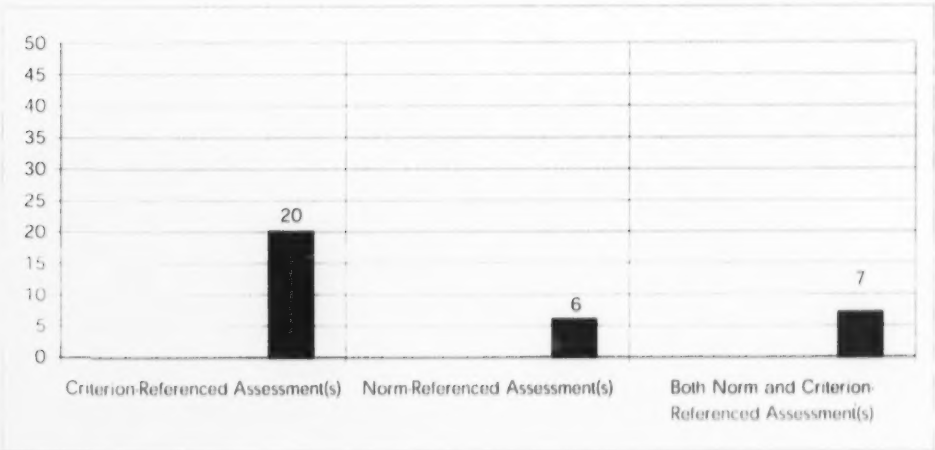
State-defined accountability systems. Thirty-three states set performance goals for schools or school districts and hold these units directly accountable for meeting these outcome goals. These states also establish rewards for meeting or exceeding state goals, sanctions for not meeting their targets, or both. Districts can exceed or supplement state accountability policies. The state performance goals vary along several dimensions, including how performance is measured and whether the performance goal is fixed or relative. Similarly, the consequences of not meeting state goals can range from school improvement planning to a loss of state accreditation or state takeover. The next four sections describe key design elements in these 33 state accountability systems.

Performance Measures for School and District Accountability

The 33 states with state-defined accountability systems use the results of state assessments as the primary measure of school and district performance. Therefore, performance measures for school and district accountability reflect the diversity of state assessment systems described earlier in this report. Twenty states use criterion-referenced assessments for accountability purposes. Six states use norm-referenced exams, such as the SAT-9, the TerraNova, or the ITBS, to measure school and district performance (Figure 8). All six states administer both criterion- and norm-referenced examinations, but have chosen to use the results of the norm-referenced tests in their accountability systems. Three of these states—California, Mississippi, and Nevada—are moving from an accountability system based on norm-referenced test results to one that incorporates results from criterion-referenced tests.

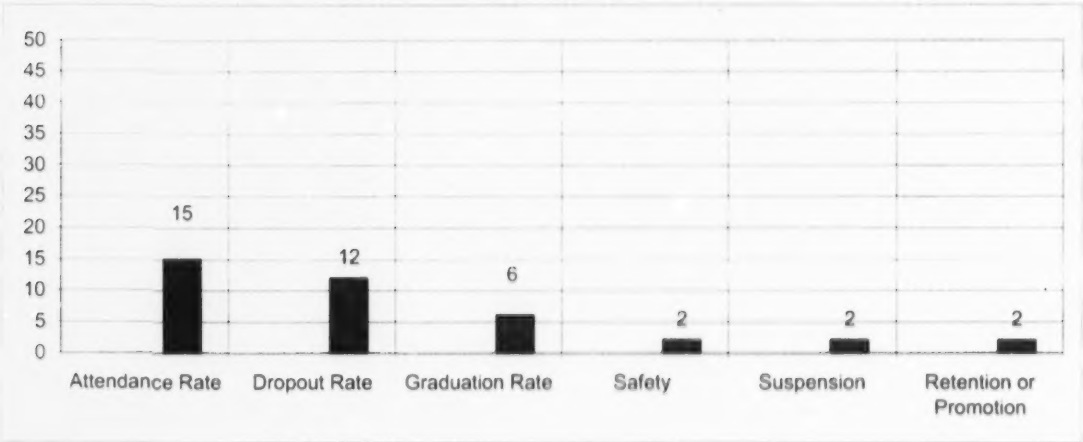
The remaining seven states use a combination of norm-referenced and criterion-referenced assessment systems as accountability indicators. Kentucky and Louisiana administer separate criterion- and norm-referenced tests and include the results of both kinds of assessments in their calculations. In the Kentucky accountability index, scores on the CTBS-5 account for five percent. In Louisiana, performance on the ITBS has a 30 percent weight for kindergarten through the eighth grade and a 20 percent weight for the ninth through twelfth grades. The Kentucky and Louisiana criterion-referenced tests are weighted 60 percent and 50 percent respectively. The other five states include a combination of norm-referenced and criterion-referenced items in their assessments.

Figure 8. Types of Assessments Used as Indicators in States with Performance-based Accountability Systems, 1999-2000



In addition to test scores, 19 of the 33 states with state-defined accountability systems use or intend to use non-cognitive indicators to measure performance of schools or districts during the 1999-2000 or 2000-2001 school years (Figure 9). The most common non-cognitive indicators are the attendance rate (15 states), dropout rate (12 states), and graduation rate (six states). Other indicators include suspension and retention rates.[6]

Figure 9. Number of States Using Non-cognitive Accountability Indicators: 1999-2000



These 19 states incorporate non-cognitive indicators in their accountability systems in one of three ways: as part of a school performance index, as discrete measures, or as a preliminary or secondary indicator. Maryland's School Performance Index, for example, is a weighted average

of a school's relative distance from satisfactory standards on the state assessment, attendance, and dropouts (high school only). Kentucky's Long-term Accountability Model includes a non-academic index with four outcomes: attendance rate, retention rate, dropout rate (middle and high school), and rate of successful transition to adult life (high school only). The non-academic index accounts for just under five percent of the total state index at the elementary level to just under 11 percent at the high school level. Similarly, Oregon's Student Behavior Rating will be based on attendance and dropout rates, with each rate converted to an index. Index measures of student performance, student behavior, and school characteristics will be combined to create an overall school performance rating.

Three states use non-cognitive variables as discrete measures. Texas gives equal weight to cognitive and non-cognitive indicators by requiring schools to meet minimum performance standards for attendance and dropout rates, as well as for state assessments. New York schools are expected to have a dropout rate below five percent, while Ohio expects its schools to achieve minimum attendance rates of 93 percent and graduation rates of 90 percent.

A few states include non-cognitive measures as a preliminary or secondary indicator or as a district option. Under the Florida A+ Plan, a school's letter grade is reduced by one level if it reports absenteeism, dropout rates, or suspension rates significantly above the state average. Colorado includes dropout, student attendance (including the number of expelled and suspended students), graduation requirements and rates, the percentage of students participating in and exempt from assessment programs, and evidence of a safe, civil learning climate in its list of indicators for district accreditation. Colorado specifies targets for student performance on the state assessment, but it leaves goals in each non-cognitive area to district discretion. While Missouri is redefining its definition of *academically deficient* schools, the state uses graduation rates of less than 65 percent as a preliminary indicator that a school district is low-performing.

Performance Goals for Schools

A key component of standards-based reform is establishing challenging standards for all students. Accountability systems are largely designed to ensure that schools and school districts make continuous and substantial progress, within an appropriate timeframe, toward the goal of all students' meeting state levels of proficient and advanced achievement. This section looks at the goals that states have established for their schools. Do they expect schools to bring all students to the proficient level, or do they have different goals against which to measure school progress?

There is wide variation in school performance goals among the 33 states with state-defined accountability systems.[7] State targets appear to vary along three dimensions: the expected level of student performance (such as basic or proficient), the percentage of students that must meet these standards, and the length of time schools have to meet their goal. Where states set their school performance goals reflects their strategy of incentives for growth and change. As we see in the next section, school performance goals interact with state definitions of adequate yearly progress, and goal-setting is in part a political process.

Most states expect to bring some or all students to proficient levels of performance. The measure of proficiency, however, is not comparable across states. States use different assessments aligned with different standards and set different cut scores for different performance levels. A student proficient on the Rhode Island assessment, for example, may (or may not) exhibit a different level or mix of knowledge and skills than a student who performs at a proficient level in Maryland or Wisconsin. A few states focus on student achievement at more basic levels of performance. Florida, for example, gives A and B grades to schools where at least half of the students reach Level 3 ("the student has partial success with the state standards") on the state assessment. Louisiana's 10-year goal is for all students to perform at the basic level. A student at this level "has demonstrated only the fundamental knowledge and skills needed for the next level of schooling."

States also differ in the percentage of students in a school that are expected to meet basic or proficient standards. Seven states specify that they expect 90 to 100 percent of students to reach proficiency, eight states specify they expect 60 to 85 percent to reach this level, and another eight states set the goal at 50 percent meeting the assessment target.[8]

Finally, states set different timelines for meeting their performance goals. Six states have established explicit target dates, ranging from five to 20 years. Examples of these targets include: 100 percent of students at standards by 2008 in Vermont or by 2010 in Oregon, and achieving a school improvement index of 100 by 2014 in Kentucky. A second group of states does not specify target dates for meeting standards, but uses progress targets as an implicit timeline for moving schools toward state performance goals. California has set 800 as an interim goal for its Academic Performance Indicator; each school has an Annual Growth Target (of at least five percent) based on the distance between its current performance and the state goal.

A few states, intending to raise their goals over time, set lower but more immediate, and in their opinion, more achievable performance goals. Texas exemplifies this strategy. When first enacting its reforms, Texas rated schools *acceptable* if 25 percent of students passed the state assessment. Having raised this threshold by five percentage points each year, Texas now finds school performance *acceptable* if 50 percent of students pass the state assessment. Virginia set a passing rate between 40 and 60 percent, depending on the subject, for the year 2000. By 2006, however, at least 70 percent of Virginia students must pass state assessments in English (except third and fifth grade, where 75 percent of students must pass) and at least 60 percent of students must pass state assessments in three other core areas (except third and fifth grade mathematics). As New York phases in its new accountability system, 90 percent of students in a school were initially expected to perform at Level 2 ("students will need extra help to meet the standards and pass the Regents exam"). Starting in September 2000, the state commissioner of education will annually set the percentage of students' performing at or above the proficient level (Level 3) schools need to meet accountability goals.

About half of the states that set performance goals have created multiple performance thresholds to distinguish low-performing schools from schools that far exceed state standards. Placement in one of the four school performance levels (exemplary, recognized, acceptable, and low-performing) in Texas is determined by performance on three indicators: the percentage of students passing the TAAS, the percentage of students dropping out of school, and the

attendance rate. Thus, for the 1999-2000 school year, Texas rated a school as low-performing if: fewer than 50 percent of all students and the students in each subgroup passed the state assessment, the school had a dropout rate higher than six percent for all students and subgroups, and the school attendance rate was below 94 percent. To obtain exemplary status, schools had to achieve a 90 percent passing rate on TAAS and a dropout rate of one percent or lower.

States also differ in the number of categories they create and the terminology they use to describe levels of student performance. Florida assigns letter grades (A, B, C, D, and F) to its five performance categories that reflect the percentage of students scoring at or above Level 2 (minimum criteria) or Level 3 (higher criteria) on the state assessment. Michigan uses accreditation terminology to classify its schools: a school receives *summary* accreditation if at least two-thirds of students score at the highest performance level on all state assessments, and an *interim* accreditation if more than half of students meet this goal on at least one assessment. Massachusetts has six performance categories (*very high* to *critically low*) and four improvement categories. Four other states—California, Kentucky, Maryland, and Vermont—assign composite index numbers to schools, showing a school's position relative to a state goal. A School Performance Index of 100 in Maryland or 800 in California, for example, means that the school has met all state standards. Under the new accountability system in Vermont, each school is notified of its Change Index Growth Target. The target is defined as the difference between a school's Baseline Index and the State Board's goal of 500 on the performance-level point scale, divided by the number of accountability cycles remaining through the 2007-2008 school year.

How Do States Define Progress?

Once states have identified performance measures and established performance goals, they must determine how they will measure annual progress toward these goals. Title I requires states to define what they consider substantial and continuous progress toward performance goals. Using these definitions of adequate yearly progress, states must identify schools and districts in need of improvement.

The 33 states with performance-based accountability systems use at least one of the three following approaches to measure school progress:

- Meet an ***absolute*** target: achieve a performance threshold or thresholds to make satisfactory progress;
- Make ***relative*** growth: meet an annual growth target that is based on each school's past performance and often reflects its distance from state goals; or
- Narrow the ***achievement gap***: reduce the number or percentage of students scoring in the lowest performance levels (Figure 10 and Table 1).

Fourteen of the states use only absolute targets as their definition of progress, while five states use only relative growth expectations. Eight states employ an absolute target and relative growth

in their definition of progress. The remaining six states use narrowing the achievement gap as at least one criterion of adequate yearly progress.[9]

Table 1. Categories of Defining School Progress in State-defined Performance-based Accountability Systems, 1999-2000

State	Meeting an Absolute Target	and/ or	Making Relative Growth	and/ or	Narrowing the Achievement Gap
Alabama	x				
Arkansas	x	Or	x		
California			x		
Colorado	x	And	x		
Connecticut	x				
Delaware [1]	x	And	x	and	x
Florida	x				
Illinois	x			and	x
Indiana	x				
Kentucky			x	and	x
Louisiana	x	Or	x		
Maryland			x		
Massachusetts [1]	x	and	x		
Michigan	x				
Missouri*			x	and	x
Mississippi*	x				
Nevada	x				
New Jersey	x				
New Mexico [3]	x				
New York [3]	x	or	x		
North Carolina	x	or	x		
Ohio*	x	or	x		
Oklahoma	x				
Oregon [2]	x				
Pennsylvania			x		
Rhode Island			x	and	x
South Carolina*			x		
Tennessee	x	or	x		
Texas	x				
Vermont [1]			x		
Virginia	x				
West Virginia	x				
Wisconsin	x	or	x	and	x

1. Planned to be implemented in 2000-2001.
2. Planned to be implemented in 2000-2001, pending Federal approval.
3. Planned to be implemented in 2000-2001, pending state board approval.

* For the purposes of Table 1, four states have been categorized based on their district accountability criteria for various reasons. Although **Missouri** holds schools as well as districts accountable, the state's achievement goals are part of the Missouri School Improvement Program at the *district* level. **Ohio** designates each *district* as Effective, Continuous Improvement, Academic Watch, or Academic Emergency and does not have school-level performance-based accountability. In two other states (Mississippi and South Carolina), the systems of accountability are in a transitional phase, moving from district to school accountability systems. The **Mississippi** state system of accreditation had ranked *districts* on a scale of 1 to 5; districts have been held harmless under this system for the 1999-2000 and 2000-2001 school years as the state moves to a system of school-based accountability. **South Carolina's** previous—and not yet entirely replaced—accountability system identified *districts* as impaired on the basis of 35 indicators: school districts were required to satisfy two-thirds of the standards for the BSAP and MAT7 achievement test results and two-thirds of the non-cognitive indicators.

Figure 10. The Number of States Using Each of the Three Methods for Defining School Progress, 1999-2000

Florida and Texas are examples of states that use ***absolute*** targets. Florida grades schools on a scale from A to F. A school earns each grade by meeting specific performance standards. For example, at least 60 percent of a school's students must score at Level 2 ("limited success at meeting state content standards") on the state assessments in reading, mathematics, and writing for a school to receive a grade of C. Schools that do not meet this criterion in *any* of the three tested areas are given a grade of F and judged as making inadequate yearly progress. Texas defines achieving the state's *acceptable* rating as adequate yearly progress. To be rated *acceptable* in 1999-2000, at least 50 percent of students in each subgroup had to pass the state assessment in reading, writing, and mathematics; the dropout rate had to be six percent or less; and student attendance had to be at least 94 percent.

The use of ***relative growth*** criteria emphasizes continuous improvement. Maryland and California have established annual goals for their schools that require continuous progress toward meeting a state-specified performance target. California recently assigned schools individualized annual growth targets that are five percent of the difference between their Academic Performance Index baseline score for July 1999 and the statewide interim performance target of 800. Maryland only requires schools to show "statistically significant" change in their School Performance Indices. The School Performance Indices, however, are re-calculated annually to reflect how far a school is from meeting state performance goals.

Eight states require schools to meet an ***absolute*** target, or make ***relative*** growth, or both. Schools in six of these states must meet an absolute target *or* make relative growth; schools in the two other states must meet an absolute target *and* make relative growth. In North Carolina, a school makes adequate yearly progress if it either meets the absolute performance minimum threshold (not more than 50 percent of students below grade level) or its expected growth goal. Under the new School Performance Rating Process, Massachusetts will require schools to meet both criteria. Each school will be assigned an overall performance rating (absolute target) and an overall improvement rating (relative growth). The state will combine these measures when placing a school in a performance category.

The remaining six states require schools to show evidence they have narrowed the *achievement gap* between low-and high-performing students. Three of these states require schools to narrow the achievement gap and to make gains on their average scores. Rhode Island requires schools to increase overall performance and increase the performance of students in the lowest-performing category by three to five percent a year. Illinois is the only state that requires schools to meet an absolute target and to narrow the performance gap. Two states use all three adequate-yearly-progress methods. Beginning in 2000-2001, Delaware will rate schools on three factors: the absolute performance of all the school's students on state assessments (*absolute performance*), the school's record in improving the performance of its students on the assessments (*improvement performance*), and the school's record in improving the performance of students at lower levels of achievements on the assessments (*distributional performance*).

Accounting for Subgroup Performance

Only six states consider the extent to which schools narrow the achievement gap between low- and high-performing students when measuring the progress of their schools, but ten states address performance differences by including subgroup performance in their accountability systems.[10] States account for subgroup performance in one of four ways:

- as a requirement for making adequate yearly progress,
- as a requirement for receiving a reward,
- as a secondary accountability indicator, or
- as a school analysis requirement.

Two states (Texas and New Mexico) include subgroup performance in their measures of adequate yearly progress. To receive an *acceptable* rating in Texas, each racial/ethnic (African American, Hispanic, White) and socio-economic (economically disadvantaged) subgroup, and the total student population in a school and district must meet the performance targets for each academic subject and non-cognitive indicator. In New Mexico, special education students, students with limited-English proficiency, and economically-disadvantaged students are identified at each school and their progress is assessed under the state accountability program. All students, including these three subgroups, are required to make adequate yearly progress.

Other states have begun to include subgroup performance in their rewards programs. Under new policies in California, Maryland, and South Carolina, state rewards and recognition will take into consideration the performance of minority students and other subpopulations in each school. To receive a grade of A or B in Florida, a school must ensure that racial/ethnic subgroups (African American, Hispanic, White, Asian, and Native American students) and poor students meet minimum performance criteria. To be eligible for rewards in Louisiana, schools are required to show improvement in the scores of their at-risk populations. New York has also proposed including subgroup performance as indicators in its reward system.

A few states use subgroup performance as a secondary indicator within their accountability systems. Arkansas considers the performance of special populations (students with disabilities, students with limited-English proficiency, and highly mobile students) in assigning points under Tier II accountability; schools may include improved performance of other subgroups as additional Tier II indicators.[11] Rhode Island schools compare performance of subgroups of their students with the statewide average performance of similar students. If a school finds a discrepancy greater than 15 percent between the achievement of its students and the statewide subgroup, it must create a plan to address the disparity. Rhode Island schools must also raise the achievement of their lowest-performing students to meet the state's adequate yearly progress requirements.

Finally, Missouri requires schools to analyze subgroup performance data. Missouri is incorporating this analysis into its accreditation process. Schools will be required to evaluate disaggregated scores to ensure that different populations are making gains at least equal to that of non-minority populations.

Sticks and Carrots: Consequences for Performance

Accountability systems create incentives for schools and school improvement by defining and measuring performance outcomes and progress, and by attaching consequences to these outcomes. Consequences for students, schools, and school districts vary across states, depending on the locus of authority (state versus local) and state willingness and capacity to intervene in low-performing schools.

Consequences for schools. Most states direct rewards and sanctions to the school level. States use a combination of public reporting, ratings, program improvement, rewards, and intervention to hold schools accountable for student performance. All 50 states have some kind of public reporting, but only 33 states impose consequences on all schools beyond reporting. The other 17 states have accountability policies that apply only to Title I schools.

All 33 states with state-defined accountability systems identify low-performing schools and have some provision for assistance, either from the state or the local school district. As discussed in greater detail later in this report, state assistance can include support in the school improvement planning process, funding for the school improvement planning process and for improvement initiatives, or technical assistance. Most of these states apply sanctions to schools that fail to improve after a specified period of time. Many states also give schools financial rewards for high levels of performance or for the improvement of student outcomes.

Consequences of Low Performance

- Mandatory public hearing
- Writing or revising a school improvement or action plan
- Mandatory technical assistance
- On-site audit or monitoring by state officials
- Probationary status or placement on state warning list
- Suspension or loss of state accreditation status
- Transfer or replacement of instructional or administrative staff at the school or district level
- Optional transfer of students
- State takeover or reconstitution
- School closure

School improvement planning. A majority of the 33 states require low-performing schools to develop improvement or action plans that identify strategies to address their areas of weakness. In addition, some states mandate public hearings to inform the public about a school's low performance and offer the larger school community an opportunity to provide input. The state may also send a team to guide the school through this planning process, provide mandatory technical assistance, and then monitor the school's improvement over the following year or years.

Intervention policies. If a school fails to meet state improvement goals within a specified time period, states may apply more intensive monitoring or intervene in the operation of the school. States like Alabama and Arkansas have developed multiple levels of state intervention. Alabama schools are identified as *Alert 1*, *Alert 2*, and *Alert 3*. An Arkansas school may go from *High Priority Status* in the first year of low performance to *Academic Distress Status I*, *II*, or *III* after a school performs poorly and does not improve after four or more years. Intervention may include state-imposed policies and practices, on-site review by state officials, increased technical assistance, or transfer or replacement of staff. Schools in states with performance-based accreditation systems may also suffer suspension or revocation of their accreditation. Students in at least six states, such as Connecticut, Florida, Louisiana, and New York, are also allowed to transfer to another school.

States may also intervene in the governance of schools. Nineteen states have enacted policies that allow them to reconstitute schools (Ziebarth, 2000). If a Colorado school that has been graded F fails to improve after two years, it could be chartered as an independent school. Although there is considerable state involvement in the process, the charter would be ultimately negotiated between the local board and the independent charter. The Maryland Board of Education has the authority to reconstitute failing schools, and recently assigned the management of three elementary schools to a private provider. A Texas school that fails to improve may be ordered to close.

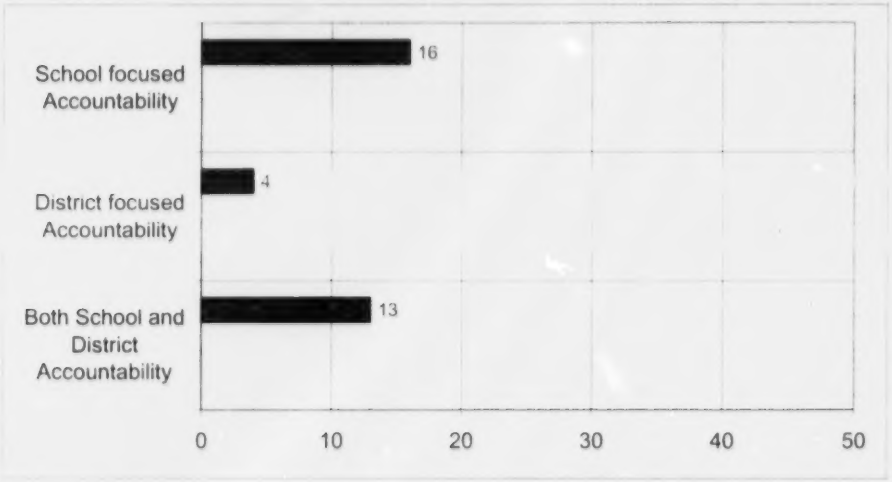
Rewarding success. Nineteen states include or plan to include monetary or non-monetary awards in their school accountability systems. The Florida legislature, for example, appropriated $15 million in 1999 for the Florida School Recognition Program. Each qualified school will be allocated up to $100 per full-time-equivalent staff member. Qualified schools include those meeting the A-grade criteria or showing *significant improvement*. Schools that improve one letter grade from one year to the next year, and F-graded schools that show significant improvement also qualify to receive additional funding. New Jersey has established an Academic Achievement Reward Program that awards $10 million annually to schools that attain *absolute success* in or

make *significant progress* toward high student achievement as measured by the state assessment system. Schools with 90 percent of students meeting state standards receive absolute-success rewards. The remaining schools are divided into five bands based on their passing rates; the 10 percent with the highest level of improvement in each band receive significant-progress rewards. A per-pupil amount is determined by dividing the $10 million appropriation by the number of students taking the test in each of the qualifying schools.

A few states recognize, but do not provide monetary rewards to, successful schools. The new Exemplary Schools Program in Massachusetts will provide opportunities for successful schools to share their knowledge with other schools in the state. At the end of each ratings cycle, schools will be eligible to apply for the program if they receive an overall improvement rating of *having exceeded expectations*, or if they receive an overall improvement rating of *having met expectations* and significantly outperformed demographically similar schools in the state in absolute performance.

Consequences for districts. States have been slower to develop district-level systems of accountability. Only 17 states hold districts accountable for student performance or for the performance of their schools (Figure 11). Four of these states—Arkansas, Mississippi, Ohio, and South Carolina—held only districts accountable in 1999-2000. Arkansas, Mississippi, and South Carolina are phasing-out their district systems and moving to school-based accountability systems.

Figure 11. Focus on Accountability: School and District Levels, 1999-2000



Thirteen states hold both schools and districts accountable for performance. Five of these states—Alabama, New Jersey, New Mexico, Tennessee, and Texas—apply the same performance criteria to schools and districts. In order to meet state standards in New Jersey, 75 percent of fourth and eighth grade students and 85 percent of high school students in each school and school district must pass the state assessment. Similarly, Texas applies the same performance requirements and levels (*low-performing* to *exemplary*) to both schools and

districts. Five other states—Delaware, Indiana, North Carolina, Rhode Island, and West Virginia—base district accountability on the percentage of schools meeting state standards. For example, a North Carolina school district in which half of the schools are identified as low-performing may lose its accreditation and suffer other consequences such as removal of the superintendent or other administrators by the local board of education. Rhode Island will designate districts that have more than 40 percent of their schools in *intervention* status as an *intervention district*. The state will send a Support and Intervention Team to work with the district in analyzing the areas that need reform. This analysis will result in a Negotiated District Agreement that details the changes to be implemented, a timeline for implementation, assistance to be provided by the state, a resource plan, the required outcomes, and indicators of successful implementation. Two states use different performance criteria for schools and districts. Pennsylvania rewards schools that show a 50-point gain in state assessment scores. Pennsylvania districts are subject to state intervention if more than half of their students score in the lowest quartile on the state test.

The other 16 states with state-defined accountability systems hold only schools accountable for performance. Some of these states require districts to develop improvement plans using school assessment results, but the districts face no consequences for low performance.

Consequences for students. As states have implemented school-based accountability systems, educators and policymakers have begun to question the lack of student incentives in these policies. Teacher success is dependent on student efforts in school, but there is nothing in a school-based accountability system that motivates students to take the tests seriously, especially in secondary schools. Nor are there any consequences for students who perform poorly on the tests. Therefore, several states and school districts have enacted promotion gates: students cannot progress to the next grade (often at transition points such as fourth grade) if they do not meet district or state performance standards. Eight states have policies for ending social promotion. California, for example, requires districts to develop standards that students must meet to be promoted at *pivotal* points, such as the third, fifth, and eighth grades. Colorado students who are not reading on grade level by the end of third grade cannot move to fourth grade reading instruction, while North Carolina requires students to pass state assessments at three *gateway* points—the third, fifth, and eighth grades. By 2008, students in 28 states will have to pass a state examination to graduate from high school. Two of these states will require that students pass either the state or a local high school assessment. In seven other states, student performance on a state assessment may be noted on a student's transcript or diploma, but passing a state test is not a requirement for high school graduation.

Alignment of General and Title I Accountability Systems Across the States

The intent of IASA was to create single and "seamless" accountability systems that would treat all schools equally. States were expected to develop aligned systems of high standards, challenging assessments and accountability, and then align their Title I programs with these policies.

We found, however, that only 22 states will have single or unitary accountability systems in place by 2000-2001. These systems hold all schools or districts to the same performance standards through the state accountability system regardless of their Title I status.[12] (See Table 2.) Some of these states, such as Florida, Kentucky, Maryland, and Texas, developed their state assessment and accountability systems prior to the enactment of IASA, and brought their Title I programs into alignment with state policies. States like Delaware, Massachusetts, New Mexico, New York, and Oregon planned to implement unitary systems of accountability for the first time in 2000-2001.

Twenty-eight states operate dual systems of accountability in which Title I or non-Title I schools are held accountable using different indicators or performance standards, or only Title I schools are held accountable by the state or district outside of the performance reporting structure.

Twelve states with dual systems of accountability have established one system of accountability for all schools and a separate system of accountability for Title I schools. Colorado and Michigan are examples of such systems. The Colorado legislature recently approved a new reporting structure that assigns letter grades to all schools based on their state assessment scores. Schools that receive a C or lower grade will be assigned an additional improvement letter grade based on change in average scores from the prior year. In contrast, Title I schools are held accountable for annual improvement on a School Index that focuses on the movement of students from the lowest to the highest proficiency levels and that sets annual performance targets over a ten-year period. In Michigan, the general accountability system places schools in one of three accreditation categories based on the percent of students who are proficient on the state assessments. Like Colorado, however, the Title I accountability system defines adequate yearly progress as narrowing the achievement gap between the highest and lowest achievement categories, not as overall performance on the state test.

Sixteen states have developed definitions of adequate yearly progress for Title I schools, but primarily hold non-Title I schools accountable either through public reporting of state or district assessment scores or through locally-defined performance measures. Arizona reports student performance on both the SAT-9 and the state criterion-referenced assessment (AIMS) at the school and district level. This is the only form of accountability for non-Title I schools. In contrast, Arizona sets annual improvement goals for Title I schools that are designed to increase the number of students scoring at the proficient level and to reduce the number of students scoring at the below-basic level on the norm-referenced test. Wyoming districts are held accountable through an accreditation process that requires periodic site visits at the district level, and all schools and districts to submit an accreditation packet to the state department of education. States like Arizona and Wyoming have a strong history of local control and have found it politically difficult to enact stronger accountability systems for all schools.

Table 2. Alignment of Title I and General State Accountability Systems, 1999-2000

State	Unitary Systems	Dual Systems
Alabama	x	
Alaska		x
Arizona		x
Arkansas		x
California	x	
Colorado		x
Connecticut	x	
Delaware [1]	x	
Florida	x	
Georgia [3]		x
Hawaii		x
Idaho		x
Illinois [3]	x	
Indiana		x
Iowa	x	
Kansas [3]		x
Kentucky	x	
Louisiana	x	
Maine [2]		x
Maryland	x	
Massachusetts [1]	x	
Michigan		x
Minnesota		x
Mississippi		x
Missouri		x
Montana		x
Nebraska		x
Nevada		x
New Hampshire [2]		x
New Jersey		x
New Mexico [3]	x	
New York [3]	x	
North Carolina	x	
North Dakota		x
Ohio	x	
Oklahoma		x
Oregon [2]	x	
Pennsylvania		x
Rhode Island	x	
South Carolina		x
South Dakota		x
Tennessee		x
Texas	x	
Utah		x
Vermont [1]	x	
Virginia	x	
Washington		x
West Virginia	x	
Wisconsin	x	
Wyoming [1]		x

1. Planned to be implemented in 2000-2001.
2. Planned to be implemented in 2000-2001, pending Federal approval.
3. Planned to be implemented in 2000-2001, pending State Board approval.

Identifying and Assisting Low-performing Schools

States' multiple accountability policies affect how they identify and assist low-performing Title I and non-Title I schools.

Identifying Low-performing Schools

In most of the unitary accountability systems, states identify schools that do not make adequate yearly progress for program improvement. Texas identifies schools and districts for program improvement if they are classified as *Unacceptable/Low Performing*. Similarly, schools in Alabama go on *academic alert* if the majority of students score below the 23rd percentile on the state's norm-referenced test. This status triggers a school-level self-study of the reasons for low student achievement and development of a school improvement plan. If a Kentucky school's accountability index falls below the *assistance line* (a line that is one standard deviation below its goal line), it is eligible for a scholastic audit to determine what kind of assistance it should receive.

Districts in a few states with unitary systems play a role in identifying schools in need of improvement. Maryland sends each of its districts a list of schools that have not made significant progress on their School Performance Indices for the last two years and recommends targeting schools with a negative change for program improvement. Maryland districts, however, apply their own criteria as well, such as mobility, relative position among schools in the county, or severity of decline. For example, one district with above-average performance identified its ten lowest-performing schools for assistance, although only two schools had been so identified by the state. In other districts, schools with a negative School Performance Index may not be identified for program improvement if indicators from other achievement data or special circumstances suggest the school should not be so identified (O'Day, 1999). Districts are primarily responsible for assisting program improvement schools, so the number of eligible schools and the severity of their need may limit the number of schools identified by district. Therefore, a school in decline may be identified for program improvement in one district, while a school with a similar accountability status might not be selected in an adjoining county.[13]

States without state-defined accountability systems generally do not identify non-Title I schools for assistance. Only Title I schools are targeted for school improvement. Minnesota, for example, does not have criteria for identifying low-performing schools or districts outside Title I. Districts are responsible for identifying and supporting schools that perform poorly on state or other tests; the extent of district involvement differs by district. In Arizona, the state has played a more active role in assisting Title I schools by creating school support teams. These teams collaborate with school-site personnel in planning and implementing schoolwide programs.

Assisting Low-performing Schools

Under IASA, districts have the primary responsibility for assisting schools that have been identified for program improvement. Such assistance includes support for school improvement planning and technical assistance in implementing these plans. States, however, must establish a

statewide system of support and improvement for Title I schools, including but not limited to those schools identified as in need of improvement. This system is supposed to include school support teams, distinguished schools, and distinguished educators.

States provide various forms of assistance to low-performing schools. The mix and level of services provided to Title I and non-Title I schools, however, varies across states and districts.

Types of assistance. We identified four kinds of assistance that states provide their schools:

Support in school improvement or corrective action planning. State departments of education provide needs assessments, on-site evaluations, assistance and training in data analysis, and other forms of technical assistance that help schools and districts to write or revise school improvement plans identifying weaknesses in student performance and strategies for improving achievement. Rhode Island's School Accountability for Learning and Teaching (SALT) initiative supports self-study, the development of school-improvement plans, school visits, and the development of a Compact for Learning that specifies what the district and the department will do to build the school's capacity.

Financial assistance for low-performing schools. Some states offer additional funding for the school improvement planning process and other school improvement initiatives. California has allocated nearly $200 million in the last two years to support the development of school action plans in over 800 schools. Each school receives a $50,000 grant that must be matched by local funds. Schools must hire external evaluators to coordinate the plan's development; these action plans also serve as funding applications for $50,000 state implementation grants. Kentucky's Commonwealth School Improvement Fund provides eligible schools with money to support teachers and administrators in developing approaches to improve instruction or management, assist in replicating successful programs developed in other districts, encourage cooperative instructional or management approaches to specific school educational problems, and encourage teachers and administrators to conduct experimental approaches to specific educational problems. Even states that have not developed special funding programs for low-performing schools may target their federal Comprehensive School Reform Demonstration program funds or other categorical aid to these schools.

Expert assistance in planning and instruction. State and local education officials or teachers are often available to provide technical assistance on best practices and other staff development at school or district sites. As discussed below, many states have developed a distinguished educator model that places experienced or retired teachers or administrators in low-performing schools. State officials are also commonly assigned to a school or region as a liaison to provide information on state standards, assessments, or accountability policies.

State- or regionally-sponsored professional development. A number of states have developed professional development programs to work in large-group settings with administrators and staff from low-performing schools. State department officials sometimes lead these training sessions; at other times, regional service centers offer the training. These sessions cover such topics as improvement planning, data analysis, best practices, and whole school reform.

Resources. States draw on multiple resources in providing assistance to schools. Many states employ ***school improvement or support teams*** composed of state officials or local education experts. Once a school or district is designated as in need of improvement in New Mexico, for example, a school support team member and liaison from the state department of education work with the school or district to improve its status. Alabama assigns school support teams to ten geographic regions of the state. Each team includes state department representatives specializing in different instructional and administrative areas. Missouri has established Success Teams composed of department personnel working with regional professional development centers. These teams work primarily in low-performing school districts to help develop plans that school staff can incorporate to improve student performance.

The use of ***distinguished or experienced educators*** to work with teachers in low-performing schools is also becoming common. Kentucky's Highly Skilled Educators program is perhaps the most widely recognized example. First implemented in 1994, this program has identified and trained over 170 Kentucky educators to work with low-performing schools in improvement planning and capacity building. Distinguished Educators in Louisiana include "highly effective educators" such as teachers, administrators, principals, university personnel, and retired teachers. They are selected and trained by the state department of education and take at least a two-year leave of absence from their current positions to help schools in *Level II* and *Level III Corrective Action*. Illinois is in the process of piloting a similar program of Educators in Residence. The Educator in Residence is a school coach who works full time in a school that is eligible for academic assistance from the Illinois Board of Education. The responsibilities and activities of an Educator in Residence vary according to the needs of the school, the individual's expertise and experience, and the relationship the Educator in Residence has with the district or school.

Staff members from state departments of education also provide services to schools and districts. An Arkansas department staff member from the Standards Assurance Unit is assigned to any district placed in *academic distress*. These specialists go to the districts on a weekly basis to monitor district improvement and provide assistance. Nevada Department of Education employees provide many of the services available to low-performing non-Title I schools. Because Indiana does not have a large cadre of outside consultants to provide local support, the state dispatches department staff, specifically from th Division of Performance-Based Accreditation. The Maryland Department of Education has extended the reach of its staff through the department's Internet site. The web site provides results of the state assessment at the school, district, and state levels; outlines the step-by-step school improvement planning process; and offers examples of specific, research-based curricular programs and reform efforts.

Many states rely on ***regional service centers and external providers*** to provide support services to schools and school districts, particularly as state departments have been downsized (Massell, 1998). Eight regional service centers located throughout Kentucky, for example, offer assistance in professional development, school and district consolidated planning, technical assistance, program design and development, and capacity building. Michigan uses a combination of intermediate units and contractors from universities and professional organizations to support low-performing schools. Ohio has divided the state into nine regions where regional coordinating teams share successful strategies and create partnerships among districts and education stakeholders. These regional coordinating teams include representatives from: colleges and

universities, the state's Families and Children First Councils, regional centers on special education and professional development, educational televisions corporations, and the Ohio Education Computer Network Sites.

System alignment and assistance. States that have not aligned their accountability and assistance programs for Title I and non-Title I schools often only provide school improvement services to Title I schools. In Colorado and Nebraska, support to low-performing schools is provided by the state Title I office. Intermediate units in Michigan and Pennsylvania include a staff member specifically assigned to provide services to all schools receiving federal funds. All schools in Georgia can receive school improvement assistance from the state, but these services are delivered based solely on voluntary self-identification by the schools.

Aligned states generally provide similar kinds of support to low-performing Title I and non-Title I schools. Some states, however, supplement general assistance programs with special services for Title I schools. In a number of cases, Title I schools receive additional financial resources or priority status in applying for grants and other funding programs. Texas Title I schools in *program improvement* are eligible for two additional resources: school support teams and the supervisory personnel responsible for overseeing the teams. Only Title I schools in Maryland participate in that state's Blue Ribbon School program which pairs high- and low-performing schools. Title I schools in Florida have priority in competitive grants, while other states make additional resource staff or distinguished educators available only to Title I schools.

Challenges in Implementing Performance-based Accountability Systems

The data presented in this report show that state responses to calls for performance-based accountability have not been uniform. State accountability systems have common elements—assessments, standards, performance reporting and, in most cases, consequences of performance, but states have found different ways to define what it means for schools to succeed, what indicators to include in their definition of success, and what the consequences will be. These variations reflect differences in state demographics, political culture, educational governance structures and policies, and educational performance.

The states and the federal government face a set of common challenges in the push for educational accountability. We end this report with a brief discussion of accountability issues, assessment issues, equity concerns, and capacity to support school improvement.

Accountability Issues

This study raises two sets of accountability issues: the differential application of accountability policies to Title I and non-Title I schools, and the limited scope of performance-based accountability systems in many states.

Supporters of Title I of the IASA hoped that this federal legislation would serve as an impetus for states to develop an integrated set of education reform policies that would apply equally to all

students and schools. Title I schools and students would be brought under the larger umbrella of state standards-based reform; states would no longer have different expectations for Title I students or different requirements for Title I schools. This vision, however, has not been realized in most states. As described in this report, more than half of the states have dual accountability systems where Title I schools are subject to different measures of adequate yearly progress. Some of these states are taking steps to create seamless school-level, performance-based accountability systems. Many states, particularly those with strong local-control cultures, will retain dual systems.

In states with dual accountability systems, adequate yearly progress requirements for Title I schools generally meet the spirit (if not the letter) of the federal legislation, although the accountability requirements for non-Title I schools may be less rigorous. This difference is not a problem *if* most or all low-performing schools participate in a state's Title I program. In North Dakota, nearly all districts have a Title I school and most districts have just one Title I school. In other states, however, it is likely that a substantial number of low-performing schools may not be subject to the more rigorous Title I accountability policies. Middle and high schools are under-represented in the Title I program, and some large, high-poverty cities are unable to serve all of their Title I-eligible schools.

Another purpose of IASA and related legislation (like Goals 2000) was to encourage states to enact accountability systems that set challenging performance goals and to hold all schools and school districts accountable for meeting these goals. Still, a third of the states do not set specific goals for their schools or do not identify low-performing schools. Although a few of these states are developing performance-based accountability systems, most rely on public reporting of student outcomes to drive school change.

Assessment Issues

State assessments are the cornerstone of state accountability systems. Policymakers face three challenges in developing assessments that are valid and politically acceptable measures of student performance. The first challenge concerns the use of norm-referenced tests to measure student performance on state standards. The majority of the 33 states with state-defined performance-based accountability systems use criterion-referenced assessments for accountability purposes, but six states rely on norm-referenced tests and seven states use a combination of norm-referenced and criterion-referenced assessments as their accountability indicators. Additional states include the results of norm-referenced assessments in their reporting systems. Some researchers claim that norm-referenced tests do not measure performance on challenging academic standards. Parents and policymakers, however, are calling for instruments that provide nationally comparative information. Also, some educators argue that norm-referenced items are aligned with some state standards, particularly at the lower grades. The issue is not whether to include norm-referenced test items in state assessment programs. The challenge is how to create an appropriate mix of norm- and criterion-referenced items, and how to determine which items are aligned with state standards and should be used to hold schools and districts accountable for student performance.

The use of multiple measures is a second area of concern for many educators and policymakers. The federal government expects states to include multiple measures in their high-stakes accountability systems, but policymakers and the education community do not have a clear or common understanding of what this means. Do multiple measures mean assessing the same content in different ways (that is, multiple measures of the same domain), assessing a range of content with multiple instruments (but possibly with one test format), assessing multiple grades in a school, or measuring non-cognitive behaviors? The U.S. Department of Education (1999) has interpreted this requirement in the first manner—meaning the inclusion of multiple approaches and formats in a state assessment system, through either one or multiple assessment instruments. Some states, however, use only a multiple-choice format in their assessment systems, while other states include open-ended or performance items, or both. A few states address the multiple-measure requirement by including more formative assessments, such as early literacy tests, in their accountability measures, while a few other states include local assessments. Some states include non-cognitive measures in their accountability systems. The use of multiple assessments with different formats and content coverage, however, can send mixed messages to teachers about what and how they should teach, and what they will be held accountable for.

Multiple measures take on even greater importance when making high-stakes decisions about individual students. In most cases, promotion and exit decisions are based on a student's performance on one test, even though this practice violates professional testing standards. The states provide few examples of using multiple assessments to make decisions related to retention or graduation. Colorado requires multiple assessments in the same domain of reading before holding a student back in that subject. Texas students have the option of passing four high school end-of-course examinations or the exit-level TAAS in order to graduate. Some argue that giving students multiple opportunities to take an exam fulfills the multiple-measures criterion, and most states with high school graduation tests do provide multiple opportunities for students to retake that assessment. Others suggest that students should be able to follow multiple paths to meeting performance standards, including course grades, projects, and portfolios, in addition to test results.

The political viability of state assessments is a third issue facing state policymakers. The high-stakes environment of student testing has sparked a public backlash against assessments in a few states. Although there is broad public support for standardized testing, few parents want to base promotion decisions entirely on the results of one test (Public Agenda, 2000). Michigan parents lobbied successfully to replace the high school exit exam with endorsements on student diplomas. The Maryland Board of Education has delayed the implementation of a more rigorous high school graduation test, fearing high failure rates and significant achievement gaps between minority and White students. Policymakers in Arizona and other states have postponed requirements that students pass their high school assessments for similar reasons.

Equity Concerns

The design of state assessment and accountability systems raises several equity issues. The first concerns the inclusion of *all* students in state and local assessment, reporting, and accountability policies. We have seen that states differ considerably in which students they test, under what

conditions, and how educators and policymakers use (or do not use) test results. The argument for including all students in these systems is strong: it promotes access to the general education curriculum and provides incentives to educate all children to higher standards. Unfortunately, large-scale assessments currently do not provide valid and comparable measures of performance for many special needs students. Test accommodations and modifications and alternate assessments do not always yield accurate and reliable information on student mastery of state standards, particularly if the content or the construct that the test is measuring is altered. Also, some testing conditions inhibit test score comparability. Much work is needed to ensure the technical and face validity of tests used to hold students and schools accountable. And we need research on alternative accountability policies for students who take tests in non-standard conditions.

In addition to having valid student performance data, schools and districts need incentives to address the educational needs of the lowest-performing students and of special needs populations. A growing number of states are making school-level data on subgroup performance readily available to educators and the public. Few states require schools to narrow the gap between the lowest- and highest-performing students as part of their accountability systems. Only two states hold schools accountable for having all groups of students meet the same performance standards.

Finally, closing the achievement gap also requires addressing inequities in opportunities to learn to high standards. Ensuring that all students have comparable learning opportunities is perhaps the most politically challenging issue that states face. As students are expected to meet more challenging standards, they need access to an academic program that addresses these standards. Students need access to teachers who have the content knowledge and pedagogical skills to teach this curriculum to a diverse group of learners. And students need access to supplemental help as they move through the system.

Capacity

Uneven access to opportunities to learn challenging academic content leads us to the fourth issue facing policymakers. Do states and districts have the capacity to support the school improvement efforts of struggling and failing schools? States and districts need knowledge, human resources, and financial resources to turn around poorly-performing schools. The optimum mix and level of resources is unknown, but states and districts report having insufficient capacity to help the number of schools that have been (or should be) identified as in need of improvement. California designated 3,144 schools as under-performing in 1999-2000, but included only 860 of these schools in the first two years of its Immediate Intervention/Under-performing Schools Program. Other states limit the number of schools they identify as low-performing to match available resources. Maryland and Connecticut are among many states that identify only their very lowest-performing schools for state assistance. Illinois policymakers recently proposed halving the number of schools identified for intervention in order to target existing state resources to just these buildings. Many states rely primarily on federal funds, particularly Title I money, to support program improvement initiatives.

Although we are learning more about how to turn around failing schools, we need considerably more research on the roles that states and districts play and on the kinds of assistance they provide to schools identified as in need of improvement under both state and Title I criteria. The U.S. Department of Education has invested heavily in a number of studies that are intended to provide extensive information about the nature and impact of instructional interventions and school improvement efforts. The findings of these studies should inform and guide the next generation of capacity-building policies.

References

Achieve, Inc. (1998). *Academic standards and assessments benchmarking evaluation for Michigan*. Cambridge, MA: Author.

American Federation of Teachers (1998). *Making standards matter 1998: An annual report on efforts to raise academic standards*. Washington, DC: Author.

American Federation of Teachers (1999). *Making standards matter 1999*. Washington, DC: Author.

Avril, T. (2000, August 23). Fourth-grade test raises scoring concerns for the second year: New Jersey students did poorly on the written part. *Philadelphia Inquirer*, p. B1.

Bond, L., Roeber, E., and Connealy, S. (1998). *Trends in state student assessment programs, 1996-97*. Washington, DC: Council of Chief State School Officers.

Council of Chief State School Officers (2000). *Profiles of state education accountability systems*. http://www.ccsso.org/introprofile.html

Council of Chief State School Officers (2000). *State accountability reports and indicators reports: Status of reports across the states—2000: Results of a 50-state survey by the CCSSO State Education Assessment Center*. www.ccsso.org/pdfs/AccountabilityReport2000.pdf

Education Week (1999, January). *Quality counts 1999: Rewarding results, punishing failure*. Bethesda, MD: Author.

Education Week (2000, January). *Quality counts 2000: Who should teach?* Bethesda, MD: Author.

Education Week on the Web at http://www.edweek.org/context/topics/issuepage.cfm?id=49.

Goertz, M.E. (2000, April). *Local accountability: The role of the district and school in monitoring policy, practice, and achievement*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Goertz, M.E., Duffy, M., and LeFloch, K.C. (2000, August). *Identifying low-performing schools: The role of Title I*. Unpublished report prepared for the U.S. Department of Education. Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

Goertz, M.E., Massell, D., and Chun, T. (1998, October). *District response to state accountability systems*. Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, New York, NY.

Massell, D. (1998). *State strategies for building capacity in education: Progress and continuing challenges.* (CPRE Research Report Series No. RR-41.) Philadelphia: University of Pennsylvania, Consortium for Policy Research in Education.

McDonnell, L.M., McLaughlin, M.J., and Morison, P. (Eds.) (1997). *Educating one and all: Students with disabilities and standards-based reform.* Washington, DC: National Academy Press.

National Center on Educational Outcomes (1999, December). *1999 state special education outcomes: A report on state activities at the end of the century.* http://www.coled.umn.edu/nceo/OnlinePubs/99StateReport.htm

O'Day, J.A. (1999). *One system or two? Title I accountability in the context of high stakes for schools in local districts and states.* Washington, DC: U.S. Department of Education.

Public Agenda (2000, October 5). Survey finds little signs of backlash against academic standards or standardized tests. http://www.publicagenda.org.

President's Advisory Commission on Educational Excellence for Hispanic Americans (2000). *Testing Hispanic students in the United States: Technical and policy issues.* Washington, DC: White House Initiative on Educational Excellence for Hispanic Americans.

Sack, J.L. (2000, June 21). Alternate-test plans prove challenging. *Education Week*, p. 1.

U.S. Department of Education (1999). *Peer reviewer guidance for evaluating evidence of final assessments under Title I of the Elementary and Secondary Education Act.* Washington, DC: Author.

U.S. Department of Education (1993). *Reinventing Chapter 1: The current Chapter 1 program and new directions. Final report of the national assessment of the Chapter 1 program.* Washington, DC: Author.

Ziebarth, T. (August 2000). "State takeovers and reconstitutions." *Policy Brief.* Denver: Education Commission of the States.

End Notes

[1] Currently, Tennessee students must pass a Competency Test to graduate from high school. Students' end-of-course test scores are part of the grades in the subjects where they are tested. The state is developing new Gateway examinations that will replace the old end-of-course examinations and become the new high school graduation requirement.

[2] This CPRE study looked at three districts in each of seven states—Colorado, Florida, Kentucky, Maryland, Michigan, Minnesota, and Texas—and two districts in California.

[3] Although states use these terms in different ways, and sometimes interchangeably, we define *accommodations* as changes in presentation, response mode, time, or setting, and *modifications* as changes altering the content of the assessment.

[4] See McDonnell, McLaughlin, and Morison (1997) for a comprehensive discussion of issues related to the inclusion of students with disabilities in state assessment and accountability systems.

[5] Many of these states have developed performance targets for their Title I schools, and some have enacted input-based accreditation policies.

[6] California and Vermont include non-cognitive indicators in their accountability policies, but did not incorporate these measures into performance calculations for the 1999-2000 school year. These states had not yet determined what weights they would assign these non-cognitive indicators.

[7] This analysis does not include separate performance goals that some of these states and the 17 other states have established for Title I schools. An earlier analysis of state Title I policies, however, found the same variation in goals (Goertz and Duffy, 2000).

[8] Four states are in the process of setting performance targets. The remaining six states do not define their long-term goals as a percentage of students meeting a proficiency standard.

[9] Twenty-two of these 33 states apply these definitions of adequate yearly progress (AYP) to all schools. The other eleven states use different definitions for their Title I and for their non-Title I schools. The AYP definitions discussed here apply to the non-Title I schools in these states.

[10] Eight of these ten states do not include narrowing the achievement gap in their measures of adequate yearly progress.

[11] Tier II indicators in Arkansas are based on trend and improvement goals for state criterion-referenced tests and school-selected indicators. Schools select five indicators where they will improve; these indicators can focus on the improving academic performance of subgroups.

[12] When categorizing state accountability systems as unitary or dual, we looked at the performance indicators, school performance goals, and measures of adequate yearly progress used to hold schools accountable and at the consequences of the accountability system. We did not include the kinds of assistance that would result from the system of accountability. Even within the category of unitary systems, we found slight differences between the indicators used to measure the performance of Title I and non-Title I schools. In West Virginia, for example, the definition of adequate yearly progress is based on performance on the SAT-9, and does not consider attendance and dropout rates which are included in the general state accountability system. As the general and Title I systems are identical with regard to what is expected of schools in terms of performance on the state assessment, we classified the state as having a unitary system.

[13] Maryland uses a similar approach for identifying reconstitution-eligible schools. Reconstitution regulations call for the identification of schools below state standards, but the state has identified only the poorest-performing schools that are in decline.